

Reconeixement de Gestos Mitjançant una Webcam

Samuel Esteban-Iglesias

Resum—Aquest document proposa un mètode de detecció de gestos amb la fase prèvia de segmentació de les mans. Per tal de realitzar la segmentació s'han combinat les Mixture of Gaussian amb el mètode Optical Flow. Posteriorment s'ha dut a terme el reconeixement de gestos emprant els descriptors de Fourier. Cadascuna de les tècniques utilitzades en el treball ha estat validada per veure si s'adequa als requisits del projecte.

Paraules clau— descriptors de Fourier, flux òptic, GMM, k-means, Lucas Kanade, pell, reconeixement, gestos, segmentació, mans.

Abstract—This paper proposes a method for detecting gestures including the previous hand segmentation phase. In order to perform this segmentation Gaussian Mixture Models have been combined with the Optical Flow method. Later, gesture recognition is done using Fourier descriptors. Each of the techniques used in this study have been validated with the intention of checking how they fit to the project requirements.

Index Terms—Fourier descriptors, Optical Flow, GMM, k-means, Lucas Kanade, skin, recognition, gesture, segmentation, hand.



1 INTRODUCTION

LA necessitat d'interactuar de noves formes amb els ordinadors ha anat creixent des de la seva aparició. Aquest fet ha suposat l'aparició de tecnologies com l'Human Interface Device (HID), en les quals mitjançant la visió per ordinador el sistema rep les instruccions directament dels usuaris a partir dels gestos realitzats. Aquest tipus de sistemes aporten l'avantatge de no requerir l'ús d'altres dispositius com poden ser el teclat o el mouse, de manera que l'usuari per si mateix pot interactuar de forma intuïtiva amb la màquina.

En l'actualitat trobem dispositius orientats al HID tals com el Kinect™ de Microsoft [1, 2], un dispositiu AdHoc que incorpora una càmera RGB i un sensor que utilitza tecnologia d'infraroigs. Aquest aporta una característica més envers les webcam que és de gran interès en la segmentació: la profunditat. El Kinect™ funciona correctament en entorns controlats d'il·luminació. En canvi si s'utilitza en exteriors, en fer servir infraroigs pot patir interferències procedents de la llum solar, fet que esdevindrà en un mal funcionament. A més a més, com no és un perifèric que s'inclougui als ordinadors no estarà tan estès com les webcam, que formen part de la majoria d'ordinadors actuals.

Aquest treball s'adreça al reconeixement de gestos mitjançant webcams. La detecció dels gestos, realitzada amb descriptors de Fourier [10, 11] requereix d'una segmentació eficient i acurada de les mans al llarg del vídeo, motiu

pel qual s'han analitzat diferents tècniques de segmentació del color de la pell (k-means [3] i Gaussian Mixture Models [4]) i el mètode d'extracció de moviment Optical Flow [7, 8], així com la seva possible combinació.

L'objectiu principal és crear un sistema detector de gestos. L'avaluació de les tècniques proposades s'ha fet sobre dos bases de dades obtingudes específicament pel desenvolupament d'aquest treball: una per validar la segmentació i l'altre per validar la detecció de gestos.

L'estructuració del document és la següent. A la secció 2 es detalla el contingut de les bases de dades. La secció 3 exposa els diferents procediments en la segmentació de les mans. La 4ª secció inclou els resultats de la validació de la hand segmentation. La 5 introdueix els descriptors de Fourier juntament amb els resultats de la seva validació. Finalment, a la 6ª secció es realitzaran les conclusions sobre el sistema obtingut.

2 DATABASE ACQUISITION

L'avaluació de les tècniques proposades s'ha fet sobre dues bases de dades expressament adquirides:

- La primera (DB1) s'utilitza en la validació de la segmentació, motiu pel qual s'han obtingut de manera particular per la posterior obtenció de resultats. Conté 3 mostres en 2 condicions de fons diferents (Fig. 1), formant un total de 6 seqüències de vídeo. Aquestes estan preses amb la millor resolució possible per tal de discriminar entre les diferents metodologies de segmentació presentades.
- La segona de les bases de dades (DB2) s'empra en la validació dels gestos. Recopila fragments amb

• E-mail de contacte: saesig@gmail.com
• Menció realitzada: Computació.
• Treball tutoritzat per: Debora Gil Resina (Ciències de la Computació)
Coen Antens (Ciències de la Computació)
• Curs 2013/14

condicions més realistes d'il·luminació, fons i resolució de vídeo per veure l'impacte d'aquests en el reconeixement de gestos (Fig. 2). Es va realitzar un "request for hands" amb la finalitat d'obtenir un total de 22 mostres en 3 tipus de fons diferents, combinats amb 2 tipus d'il·luminació. Formen un total de 132 vídeos.

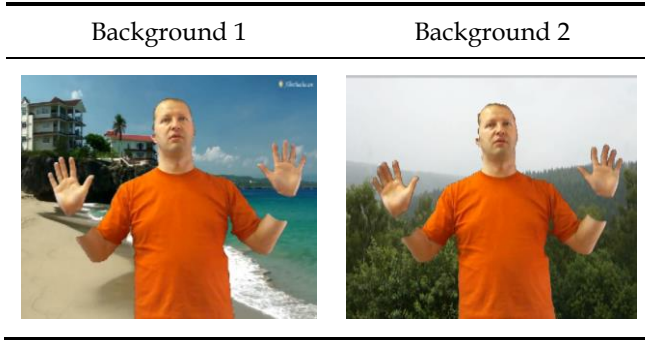


Fig. 1. Mostra 1 (DB1) amb els dos tipus de fons.

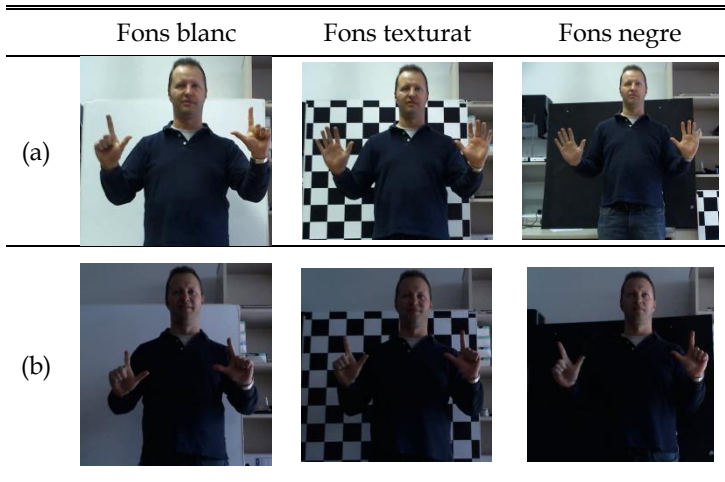


Fig. 2. Mostra 1 (DB2) en les diferents condicions d'adquisició (a) bona il·luminació. (b) il·luminació pobre.

3 HAND SEGMENTATION WORKFLOW

A l'hora de segmentar una imatge en regions d'interès (ROI) cal decidir quines característiques s'utilitzaran per fer-ho. En aquest treball s'ha optat per seleccionar l'aparença (color) de la pell i el moviment ja que són dues característiques que defineixen les mans i que les webcam ens permeten extreure.

El workflow de la hand segmentation consisteix en, a partir d'un fotograma de vídeo, realitzar la segmentació per color i la segmentació per estimació del moviment. Una vegada es realitzen es finalitza amb l'operació lògica AND entre totes dues segmentacions per combinar ambdues característiques (Fig. 3).

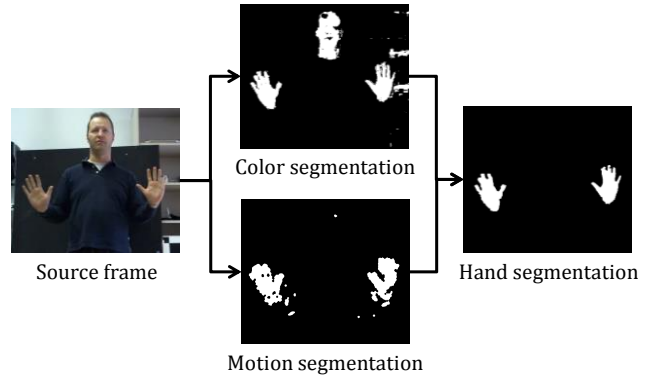


Fig. 3. Hand Segmentation Workflow.

3.1 Skin detection

En el workflow del skin detection (Fig. 4) primer es detecta la cara per generar la classe de pell en l'espai de color HSV. A continuació s'assignen els píxels de pell per tota la imatge en funció de la classe generada al pas anterior. Finalment, s'apliquen les Geometric-Spatial constraints.

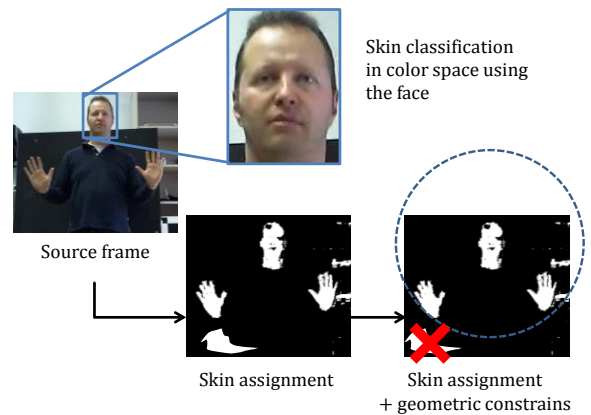


Fig. 4. Skin Detection Workflow.

Vist el workflow d'aquesta secció de l'aplicació es passa a explicar en detall cadascuna de les fases.

Una de les característiques seleccionades per dur a terme la segmentació ha estat el color de la pell. Es tracta d'un atribut que en funció de l'espai de color que es faci servir ens facilitarà la posterior segmentació. Per tal de trobar el que més s'adequa a les necessitats del projecte s'han provat els espais HSV (Hue Saturation Value) i RGB sobre la imatge (a) de la Fig. 5, resultats amb els quals es podrà conèixer la ubicació dels píxels de pell. Es prova amb l'espai HSV pel fet que l'article [6] indica, a la secció "B. Skin-color segmentation" que aquest s'utilitza per segmentar regions de pell en imatges a color.

D'una banda, a l'espai RGB els píxels es distribueixen generant una forma ovalada que travessa gran part del cub (des del punt (60,40,40) fins al (180,160,160)). D'altra banda, a l'espai HSV els píxels acostumen a estar ubicats en els valors d'H més petits que 0.3. Els valors d'H variaran en funció de la tonalitat de la pell de l'usuari però sempre es trobaran propers a l'indicat.

A partir d'aquests resultats es considera més conveni-

ent fer servir HSV, ja que la distribució de píxels obtinguda és més adequada a l'hora de fer el clustering de la pell. Cal també tenir en compte que en aquest espai és aconsellable prescindir del canal V [6], ja que es correspon amb el valor d'intensitat de llum: volem treballar amb la crominància, no amb intensitats que ens facin dependre de la il·luminació.

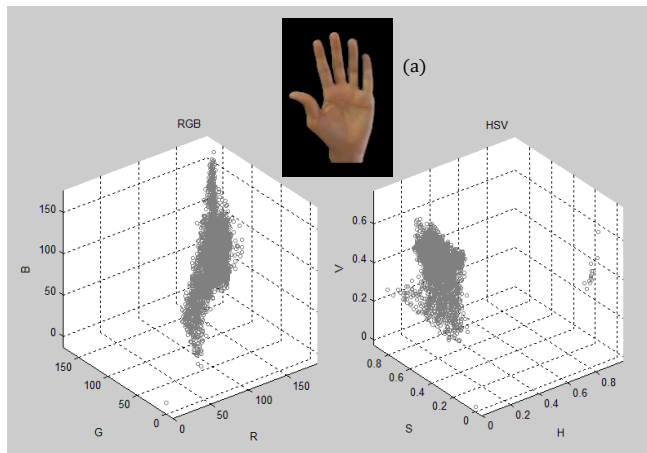


Fig. 5. Espais HSV i RGB generats a partir de (a).

La classificació dels píxels de la imatge es pot fer tant de forma supervisada com de no supervisada. Tot i això, la estratègia supervisada faria que s'excloués una de les característiques interessants que es vol incloure en el projecte: no descartar la variabilitat entre usuaris pel seu color de pell. Per aquest motiu, com es vol que el sistema s'adapti a les característiques de cada usuari s'utilitzarà una estratègia no supervisada. La classe pell en el núvol de punts es definirà a partir de la clusterització donada per la regió de la cara, ja que suposem que és la regió amb més píxels de pell a la imatge que té menys variabilitat entre adquisicions (perquè no inclou el fons).

Pel que fa a la obtenció de la regió de la cara s'ha emprant l'algorisme Viola & Jones [9].

I. Face Unsupervised Clustering

Per la clusterització de la regió facial s'han provat dos algorismes de classificació, un de no paramètric, K-means i l'altre de paramètric, Gaussian Mixture Models (GMM). En ambdós casos el kernel/gaussiana que representarà la pell serà el que contingui més observacions ja que suposem que a la cara hi ha més píxels de pell que de no pell. A continuació s'aprofundeix en el funcionament de cadascun d'ells.

K-means

K-means [3] és un algorisme no paramètric que divideix en k grups un conjunt de n observacions: els píxels de la imatge (observacions) es distribuïran formant k agrupacions en funció del seu color. Determinar k és difícil, ja que depèn de la variabilitat en la imatge (nombre d'objectes, irregularitat del fons, etc). Per aquest motiu s'utilitza la regió de la cara, que permet cercar un valor k

més acotat. Per trobar la k adient s'ha aplicat el k-means sobre les mostres de la segona base de dades, ja que incorpora més variabilitat de mostres.

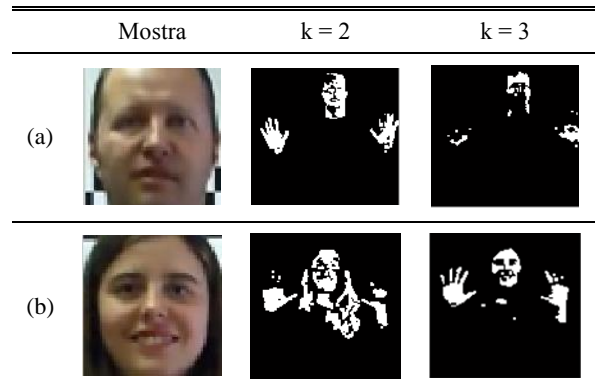


Fig. 6. Mostres (DB2) segmentades a partir del k-means de la cara.

El valor k ideal es troba entre 2 i 3: veiem com les mostres amb cabell ros acostumen a millorar els resultats amb $k = 3$ (Fig. 6 (b)), mentre que la resta de mostres acostumen a empitjorar (Fig. 6 (a)). L'explicació a aquest fenomen és que els valors H i S del cabell ros són molt similars als de la pell (centroïdes blau i verd de Fig. 7): si es fan servir dos kernels el color del cabell quedarà inclòs en el kernel de pell. Amb tres kernels l'algorisme aconsegueix separar el cabell ros de la pell.

GMM

Com s'indica a la introducció de [5], el GMM és una funció de densitat de probabilitat paramètrica representada com la suma dels pesos de les components de densitat de les diverses gaussianes que el formen: es tracta de n gaussianes que emprant l'Expectation Maximization s'adapten a un conjunt d'observacions de forma paramètrica (Fig. 7) [4]. Acostumen a utilitzar-se en la clusterització de dades, motiu pel qual és interessant aplicar-les en aquest projecte. De la mateixa manera que cal especificar el nombre de clústers en el k-means, s'especificarà el nombre de gaussianes per la GMM sobre la cara. Per trobar el nombre adient s'ha aplicat sobre les mostres de la segona base de dades, pel fet que incorporen més variabilitat entre les mostres.

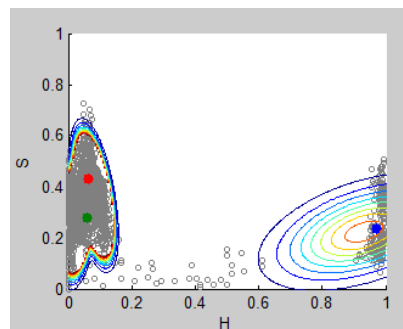


Fig. 7. Distribució de píxels, kernels i gaussianes per la mostra (b) amb $k=3$.

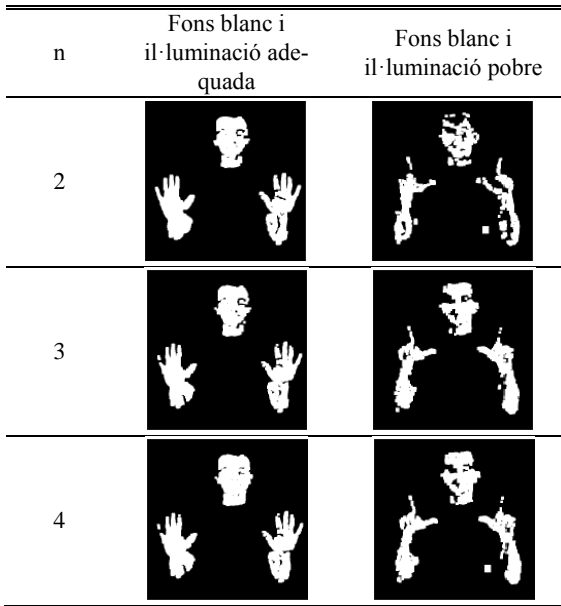


Fig. 8. Mostra 22 (DB2) segmentada segons la GMM.

Els algorismes no paramètrics com el k-means tenen l'avantatge d'adaptar-se més als clústers de les observacions encara que a l'hora de classificar presenten inconvenients, com per exemple les assignacions de classes del k-nn a les fronteres. En canvi, els paramètrics s'ajusten als clústers no amb tanta precisió com els no paramètrics però amb l'avantatge de tenir una classificació més realista.

II. Hand Classification

Una vegada es disposa de les dades de la cara clusteritzada es pot procedir amb la classificació de la resta de la imatge. D'una banda, la classificació a partir dels clústers del k-means s'ha realitzat amb els k-nn (mètode 1) i també amb la distància euclidià als centroides (mètode 2).

D'altra banda, la classificació dels píxels a partir de la clusterització de la GMM es duu a terme mitjançant la Probability density function (PDF) de la Gaussian Mixture Distribution (mètode 3). La pell es veurà representada pel 98.08% de la població ($1 - \log(0.04)$), mentre que la resta seran els píxels classificats com no pell.

III. Geometric-Spatial Constrains

A la sortida de la hand classification cal aplicar diverses restriccions referents a la coherència espacial, ja que es volen detectar només les regions de les mans. Amb aquest objectiu:

- Com les mans sempre es troben properes al cap s'eliminen els blobs que es troben a una distància del cap superior a 4 caps (Fig. 9 (a)).
- Tots els blobs amb una mida inferior al 25% del cap s'eliminen (Fig. 9 (b)).

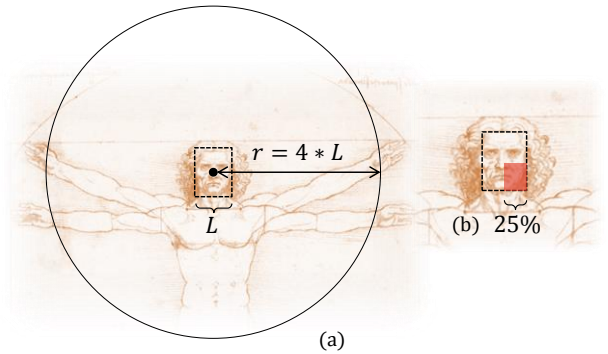


Fig. 9. (a) Restriccions geomètriques (b) Mida mínima

Una vegada explicada la segmentació per color de pell es passa a parlar de la segmentació per moviment.

3.2 Motion Estimation Segmentation

La sortida de la skin detection ofereix resultats similars al de la Fig. 10. Es pot apreciar que no tots els píxels detectats com pell ho són realment: cal tenir en compte que a la imatge trobarem objectes amb una tonalitat similar. Com aquests objectes que formen part del fons romanen immòbils durant tota la seqüència de vídeo i les mans de l'usuari no, atorgarem al sistema la capacitat d'eliminar els objectes similars a la pell si tenim en compte el moviment dels píxels entre els diferents frames.

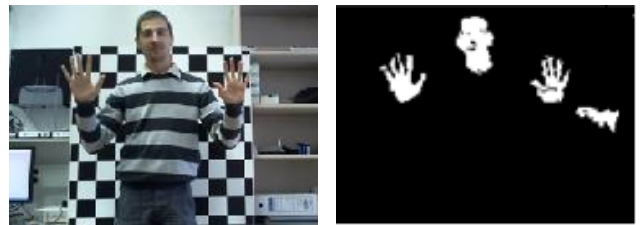


Fig. 10. Mostra 3 segmentada per color

A la introducció de [8] es descriu el Optical Flow com la distribució de les velocitats aparents del moviment segons els patrons de brillantor d'una imatge. Hi ha diversos mètodes per dur a terme l'OF, uns de locals, com Lucas Kanade i d'altres globals com Horn Schunck, entre d'altres. En el treball s'utilitzarà el primer d'aquests dos degut a que és paral·lelitzable.

Lucas Kanade

Lucas Kanade és un mètode local de càlcul del Optical Flow [7]. Fa servir un kernel a cada punt de la imatge en lloc d'utilitzar el gradient de la imatge sencera, com fan mètodes com el Horn Schunck. Aquest kernel (amb una mida de finestra que especifica la quantitat de píxels que aquest inclou) barreja el moviment dels píxels mitjançant una gaussiana amb la finalitat de determinar el moviment en aquella regió de la imatge. Com es tracta d'un mètode que treballa localment es pot convertir en un procés paral·lelitzable, de manera que mitjançant diversos fils

d'execució la tasca es durà a terme més eficientment que amb un mètode global que no pot ser paral·lelitzable.

Un dels paràmetres ajustables de Lucas Kanade és la mida de la finestra: com més gran més es suavitzarà el moviment (s'inclouen més píxels als kernels) i com més petita més soroll tindrem, ja que es perd part de l'smooth.

Hi ha d'haver un trade off entre el moviment de l'usuari que es vol capturar i el soroll en la imatge: si es volen detectar moviments amb magnitud petita s'obindrà més soroll.

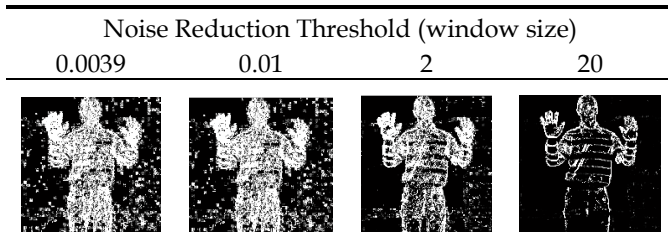


Fig. 11. Lucas Kanade sobre la mostra 3

3.3 Combination

Una vegada es disposa de la imatge segmentada per color i la segmentada per moviment es pot realitzar l'operació lògica AND entre totes dues amb l'objectiu d'obtenir els beneficis de cadascuna (Fig. 12).

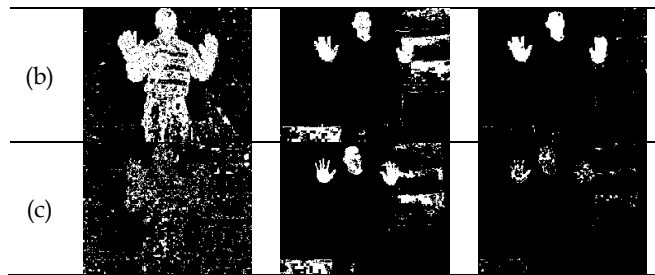
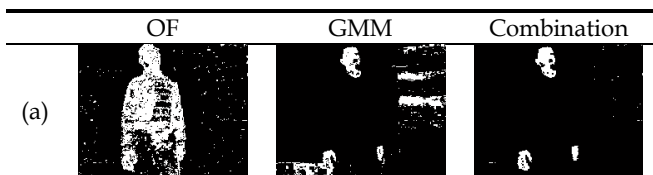


Fig. 12. Combinació skin segmentation i OF sobre la mostra 3

Les dues primeres fileres de la Fig. 12 mostren el correcte funcionament de l'eliminació d'objectes de fons mentre que la tercera filera no: (a) i (b) s'han capturat en instants de temps on l'usuari ha estat movent-se mentre que a (c) l'usuari ha estat força quiet. Això implica que si hi ha poc moviment a la imatge el resultat final de la combinació quedarà embrutat. Per tal de solucionar aquest problema es tindrà en compte només la segmentació per color de pell quan el moviment en la imatge sigui reduït.

4 HAND SEGMENTATION VALIDATION

La Fig. 13 mostra com queden els diferents mètodes explicats prèviament (sense aplicar les Geometric-Spatial Constrains) en diferents condicions de fons i il·luminació de les mostres de la DB2. Només es mostra de forma explicativa, ja que els resultats, com s'ha indicat a la secció 2 s'obtenen a partir de la DB1.

La Fig. 14 indica l'accuracy de cadascun dels mètodes per les mostres de la DB1 de dades amb les Geometric-Spatial Constrains. L'accuracy es calcula utilitzant els resultats de la matriu de confusió de la següent manera:

$$Accuracy = \left(\frac{TP}{TP + FN} + \frac{TN}{FP + TN} \right) / 2$$

Mètode	Fons blanc +llum correcta	Fons negre +llum correcta	Fons texturat +llum correcta	Fons blanc +llum feble	Fons negre +llum feble	Fons texturat +llum feble
1 k=3						
2 k = 3 th=0.2						
3 n = 3						

Fig. 13. Mostra segmentada en diferents condicions d'entorn amb cada tècnica

Accuracy GMM+PDF (mètode 1)			Accuracy k-means+knn (mètode 2)			Accuracy k-means+euclidia (mètode 3)		
Sample	Background		Sample	Background		Sample	Background	
	1	2		1	2		1	2
1	95,54%	95,65%	1	70,75%	68,52%	1	74,87%	77,04%
2	95,56%	96,79%	2	66,80%	71,64%	2	73,37%	70,74%
3	95,61%	95,48%	3	67,84%	64,08%	3	67,50%	64,82%

Fig. 14. Validació de la hand segmentation

Els resultats d'accuracy obtinguts són molt positius pel que fa al mètode 1, mentre que per la resta els valors empitjoren: en el cas del mètode 2 trobem el problema del k-rn als edges dels kernels i en el tercer mètode el fet d'utilitzar un radi al centroid que no s'adapta a la forma del kernel.

5 GESTURE RECOGNITION

Fins aquest punt s'ha tractat la part del treball corresponent a la segmentació de les mans. Una vegada obtingudes les regions d'interès es pot procedir amb la detecció de gestos.

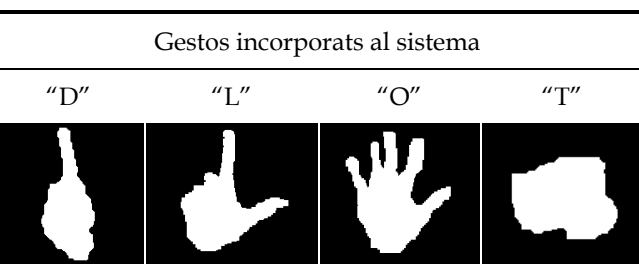


Fig. 15. Exemples de mostres de les quatre classes

Els descriptors de Fourier [10, 11] s'utilitzen en el reconeixement de formes, motiu pel qual és pràctic emprar-lo en el reconeixement de gestos. Entre les seves característiques es troba la invariància a la rotació i a l'escalatge. El procediment que es realitza per tal d'obtenir aquests descriptors és el següent:

1. Es fa el subsampling de la imatge amb el gest de la mà desitjada, obtenint un vector de punts.
2. A aquest subsampling s'aplica la Fast Fourier Transform (FFT), de manera que s'aconsegueix un vector que descriu la forma mitjançant freqüències.
3. Aquest vector és posteriorment utilitzat per la detecció de gestos mitjançant un mètode de classificació com pot ser k-nn.

El subsampling de les formes consisteix en recórrer la silueta de la imatge i extreure'n 64 punts. Aquest subsampling, una vegada s'aplica la FFT es converteix en un vector on cada posició representa una freqüència de la forma: les freqüències baixes es corresponen amb el con-

torn de la forma, mentre que les altes amb els detalls. En funció de les freqüències que s'utilitzin per definir les classes estarem seleccionant més o menys detall en les formes.

Primer de tot cal conèixer si els contorns de les classes són prou diferents entre elles com per ser reconegudes pel sistema. Per descobrir-ho es poden representar gràficament les característiques de cadascuna, amb l'objectiu d'apreciar les similituds i diferències (Fig. 16). La divergència entre les línies ens garanteix que el sistema disposarà de la suficient informació com per poder diferenciar-les. Les dues úniques classes que tenen certa similitud són la "L" i la "mà oberta".

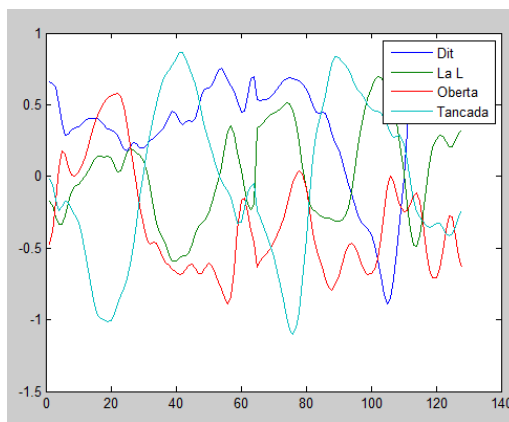


Fig. 16. Característiques de cada classe

Una vegada sabem que es poden reconèixer, hem de decidir la quantitat de freqüències del vector de característiques que empremerem.

El nombre a escollir, situat entre un mínim de 2 i un màxim de 64 es pot seleccionar comparant els diferents gestos per diferents quantitats, mirant si el valor escollit és suficient o no per distingir-les. A la Fig. 17 es mostra una comparativa entre les 4 classes del treball i diversos valors de freqüències. Es pot apreciar com la quantitat suficient es troba al voltant de 32: per sota algunes classes com la mà oberta poden quedar distorsionades mentre que afegir-ne més no n'aporta res, en tot cas, complexitat computacional.

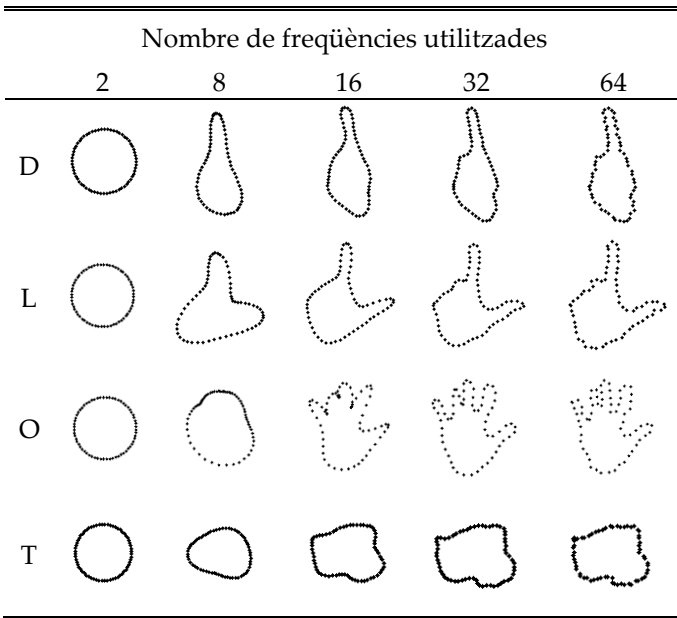


Fig. 17. Reconstrucció d'imatges segons el nombre de freqüències

La classificació en aquest cas és supervisada, ja que es requereix d'un conjunt d'imatges per la posterior classificació. Aquestes imatges han estat extretes de la segmentació realitzada a la segona base de dades.

Validation

A l'hora de validar la detecció de gestos s'ha dividit el conjunt de 22 mostres de la DB2 en dos parts:

- El conjunt de train s'obté carregant les 11 primeres mostres per gest (44 imatges en total) i generant els descriptors promig per cada classe. D'aquesta manera tindrem quatre vectors promig (un per cada classe).
- El conjunt de test, sobre els quals es realitzarà la detecció de gestos estarà format per la resta d'11 mostres.

Per cada mostra s'han seleccionat 25 frames separats en el temps per evitar la similitud entre fotogrames adjacents.

Gesture detection		Prediction			
		Dit	L	Oberta	Tancada
Actual	Dit	72,34%	8,51%	4,26%	14,89%
	L	10,26%	61,54%	2,56%	25,64%
	Oberta	13,04%	43,48%	30,43%	13,04%
	Tancada	20,00%	14,00%	8,00%	58,00%

Fig. 18. Resultats de la validació del reconeixement de gestos per la DB2

Els resultats obtinguts no són gaire bons: les classes més encertades són les del dit i la L (72,34% i 61,54%). Aquest resultats, però, són força baixos. La resta de classes no superen el 60%, i en el cas de la mà oberta el per-

centatge d'encert és inferior al de la confusió amb la L. Aquests resultats són fruit de la resolució de la imatge: a més de que la segmentació disposa d'una baixa resolució per classificar la imatge, les operacions morfològiques provoquen que les formes originals quedin modificades degut a l'opening i al closing (Fig. 22). Cal recordar també la similitud entre les classes "mà oberta" i la "L" observades a la secció 5, element que també pot estar perjudicant. A l'apèndix A2 es poden veure resultats de la segmentació amb la detecció aplicada.

A conseqüència d'aquests resultats s'ha optat per realitzar la validació de la mateixa manera amb la DB1, ja que disposa d'una millor resolució en les imatges.

Gesture detection		Prediction			
		Dit	L	Oberta	Tancada
Actual	Dit	88,10%	0,00%	7,14%	4,76%
	L	0,00%	96,67%	3,33%	0,00%
	Oberta	0,00%	0,00%	92,22%	7,78%
	Tancada	6,67%	2,38%	43,81%	47,14%

Fig. 19. Resultats de la validació del reconeixement de gestos per la DB1

Els resultats en el cas de la Fig. 19 són millors respecte a la DB2 excepte per la classe de la mà tancada, on es confon força amb la mà oberta. Veiem doncs que la resolució de l'adquisició de les imatges té força impacte en la detecció de gestos. A l'apèndix A1 es poden veure resultats d'aquesta detecció.

Els resultats que s'obtenen en la classe de la mà tancada es produeixen pel fet que les diferents mostres realitzen aquest gest de forma diversa, introduint una variabilitat en el sistema que no pot ser detectada eficaçment.

6 CONCLUSIONS

S'ha presentat l'estudi de diversos mètodes de segmentació de mans emprant dos característiques: el color de pell i el moviment. Cadascuna de les tècniques utilitzades en la classificació de les mans ha estat validada mitjançant la primera base de dades. Dels resultats obtinguts es pot dir que utilitzar la GMM i classificant amb la PDF proporciona una bona relació d'encerts, mentre que utilitzar el k-means+knn (que té problemes als edges) i el k-means+distància als centroids (que els veu perjudicat per només considerar el radi del centroid i no la forma real del clúster) l'empijora.

A més a més de la segmentació de les mans, s'ha aplicat una tècnica de detecció de gestos utilitzant la segmentació de les mostres de la DB2, dividint-la per la meitat entre train i test. Els resultats obtinguts amb aquesta han estat força baixos, de manera que s'ha validat també amb la DB1. Veiem com en la segona validació els resultats milloren notablement per la majoria de classes, indicant que la resolució de les imatges té cert impacte en la detecció de gestos proposada.

L'objectiu de crear un sistema detector de gestos ha estat assolit satisfactòriament. D'altra banda, inicialment es

va pensar en portar la versió desenvolupada de Matlab a OpenCV però finalment s'ha considerat que queda fora de l'àmbit del projecte, ja que aquest es centra en el desenvolupament de les tècniques de segmentació i detecció dels gestos, no en la creació del sistema a un altre llenguatge.

Com a future work es pot realitzar la portabilitat anteriorment esmentada, estudiar l'impacte del Optical Flow sobre la segmentació per color de pell i comprovar l'efecte d'aplicar altres tècniques de Optical Flow en la segmentació.

AGRAÏMENTS

M'agradaria agrair a la Dèbora Gil i en Coen Antens la acceptació per portar endavant el meu projecte i el fet d'haver-me guiat al llarg del seu desenvolupament en els moments en que més ho he necessitat. Voldria també anomenar en Jordi Abella per la informació dels descriptors de Fourier proporcionada i la Patricia Márquez per introduir-me en el món del Optical Flow. Finalment, a tothom qui va participar en la request for hands amb la qual vaig poder realitzar la DB2 i a la resta de gent del CVC, gracies.

REFERÈNCIES

- [1] Z. Ren J. Meng J. Yuan et al. "Robust Hand Gesture Recognition with Kinect Sensor", 2011
- [2] N. Silberman, R. Fergus, "Indoor Scene Segmentation using a Structured Light Sensor", U.S.A, 2011
- [3] T. Kanungo, D. M. Mount, N. S. Netanyahu et al, "An Efficient k-Means Clustering Algorithm: Analysis and Implementation", USA, 2002
- [4] MathWorks "Gaussian Mixture Models" Accedit a l'Abril , 2014. [Online] Disponible: <http://www.mathworks.es/es/help/stats/gaussian-mixture-models.html>
- [5] D. Reynolds, "Gaussian mixture models", USA, 2009
- [6] P. Prasertsakul, T. Kondo, T. Phatrapornnant et al. "A Robust Hand Segmentation Method Based on Color and Background Subtraction"
- [7] P. M. Valle, "Conditions Ensuring Accuracy of Local Optical Flow Schemes", UAB, 2010
- [8] Berthold K.P. Horn, Brian G. Schunck. "Determining Optical Flow", U.S.A, 1981
- [9] P. Viola, M. Jones, "Robust Real-time Object Detection", Canada, 2001
- [10] S. Conseil, S. Bourennane and L. Martin. "Comparison of Fourier Descriptors and Hu Moments for Hand Posture Recognition", 2007
- [11] J. Abella, C. Antens et al. "Multi-modal Descriptors for Multi-class Hand Pose Recognition in Human Computer Interaction Systems", 2013

APÈNDIX

A1. DETECCIÓ DE GESTOS DB1

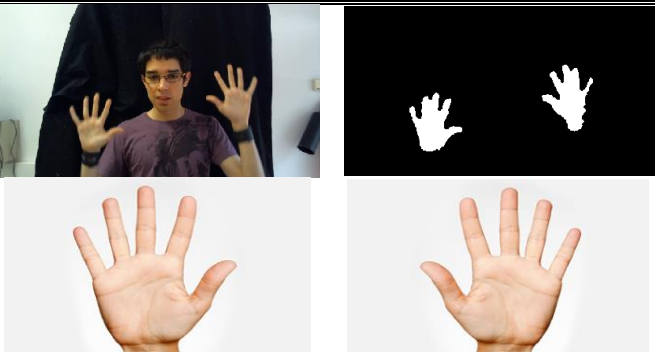


Fig. 20. Mostra 3 (DB1) amb gestos detectats

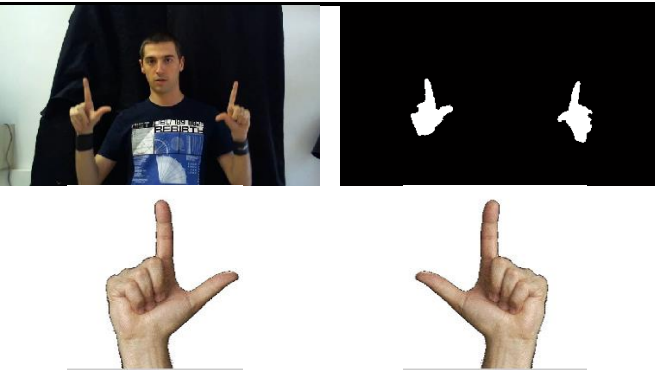


Fig. 21. Mostra 2 (DB1) amb gestos detectats

A2. DETECCIÓ DE GESTOS DB2



Fig. 22. Mostra 20 (DB2) amb gestos detectats.

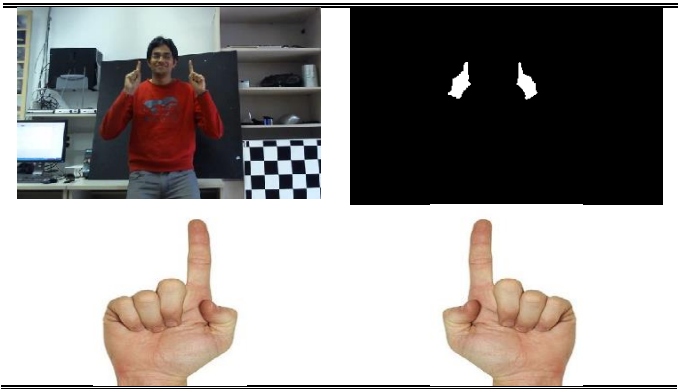


Fig. 23. Mostra 12 (DB2) amb gestos detectats.



Fig. 24. Mostra 9 (DB2) amb gestos detectats.

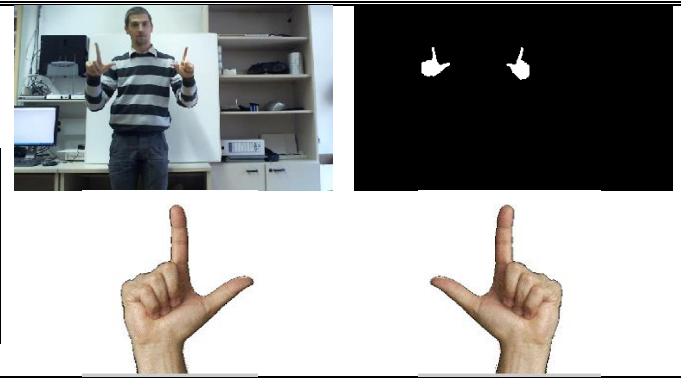


Fig. 25. Mostra 3 (DB2) amb gestos detectats.



Fig. 26. Mostra 11 (DB2) amb gestos detectats.